

Percolation on trees as a Brownian excursion: from Gaussian to Kolmogorov-Smirnov to Exponential statistics

Francesc Font-Clos¹ and Nicholas R. Moloney²

¹*ISI Foundation, Via Alassio 11/c, 10126 Torino, Italy*

²*London Mathematical Laboratory, 14 Buckingham Street, London WC2N 6DF, UK*

(Dated: June 14, 2016)

We calculate the distribution of the size of the percolating cluster on a tree in the subcritical, critical and supercritical phase. We do this by exploiting a mapping between continuum trees and Brownian excursions, and arrive at a diffusion equation with suitable boundary conditions. The exact solution to this equation can be conveniently represented as a characteristic function, from which the following distributions are clearly visible: Gaussian (subcritical), Kolmogorov-Smirnov (critical) and exponential (supercritical). In this way we provide an intuitive explanation for the result reported in R. Botet and M. Płoszajczak, *Phys. Rev. Lett* **95**, 185702 (2005) for critical percolation.

INTRODUCTION

Mappings that connect seemingly unrelated models are regularly used in statistical physics. They not only facilitate calculations but provide additional intuition about the original models. Well-known examples include the Ising model interpretation for a lattice gas[1], the bosonic interpretation for the partitioning of an integer into summands[2], or the Coulomb gas interpretation for the eigenvalues of certain random matrices[3]. In practice, the coincidence in distribution of an observable in one model may spur the search for a mapping to its counterpart observable in another model. In this Letter, we adopt this route inspired by the Kolmogorov-Smirnov (KS) distribution, which has been noticed in a variety of contexts. It is the distribution of a test statistic for comparing between empirical and theoretical distribution functions, and is well understood to describe the absolute maximum value of a Brownian bridge[4]. With this insight, it can be related to other Brownian observables[5]. More surprisingly, it is the distribution of the integrated mean-squared fluctuations of a periodic Brownian signal[6], and therefore also accounts for the roughness of a 1d periodic Edwards-Wilkinson interface in the steady state[7]. It also describes the sizes of clusters in a mean-field aggregation process[8].

This paper focuses on the finite-size scaled distribution of percolating cluster sizes on a Bethe lattice at the critical point, first computed by Botet and Płoszajczak[8, 9] to be the KS distribution. We demonstrate that there is indeed a connection to Brownian motions, thereby obtaining a more intuitive understanding of their result. In this way, we demystify the coincidence in distribution and bring the result into the existing fold of knowledge about Brownian motions and associated observables[5, 10].

SETTING UP THE PROBLEM

We consider site percolation on a finite Bethe lattice of size L , coordination number z and site occupation probability p , with critical occupation probability $p_c = 1/(z-1)$ [11]. There is a distinguishable site at the center, called the root, from which distances or ‘heights’ to other sites can be measured. With a subsequent mapping to Brownian excursions in mind, it is convenient to define the root to be at $h = 1$. Neighboring sites in the first generation are then at $h = 2$, sites at the boundary at $h = L$, etc. We say that the system percolates if there is at least one path of occupied sites from the root to any boundary site. The percolating cluster, containing the root, can be thought of as a rooted tree of fixed height. The size S of the percolating cluster is the number of sites forming the cluster that contains the root (which may include more than one path to the boundary). If the system does not percolate, we set $S = 0$ for convenience. In what follows, we will be concerned with the distribution of the size of the percolating cluster *given that* the system percolates,

$$\mathcal{P}(s) \equiv \text{Prob}[S = s : S > 0]. \quad (1)$$

In particular, we will direct our attention to the limit of large system sizes $L \gg 1$. A well-normalized limiting distribution for the rescaled percolating cluster size will therefore depend on how S scales with L for different values of p .

MAPPING TO BROWNIAN EXCURSION

The well-known *depth-first search*[12] (also known as a Harris walk) gives a bijection between rooted trees and excursions. Here we make use of an asymptotic version of this technique for $L \gg 1$ to map the percolation problem to a Brownian excursion with specific boundary conditions.

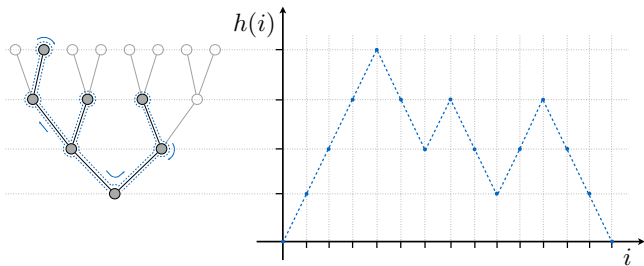


FIG. 1. The correspondence between trees and positive paths in a system of size $L = 4$. The percolating cluster has size $S = 7$, and the associated walk has duration $T = 2S = 14$. Left: The percolating cluster, traversed in depth-first search order, as indicated by the arrows next to the edges. Right: The associated path, with an initial and final step appended so that the walk does not return to $h = 0$ before the entire tree has been traversed.

Consider a realization of the percolating cluster, traversed in depth-first search order starting from the root at time $i = 0$, see Fig. 1. Each *edge* is traversed exactly twice, once in each direction, and the pairs $(i, h(i))$ form a positive walk of length $2S$. To have a well-defined excursion that terminates at $h = 0$ once the tree has been traversed and not before, an up-step and a down-step is appended to the beginning and end of the path. For large L , the walk $(i, h(i))$ can be approximated by a Brownian excursion. If we further think of percolation on the Bethe lattice as a Galton-Watson branching process with binomial offspring distribution $X \sim \mathcal{B}(z-1, p)$, then the drift v , diffusion constant D and initial condition x_0 of the associated excursion can be determined simply by inspecting and matching some well-known results. For example, the probability for an unbiased excursion to reach level k before 0, starting from x_0 , is x_0/k (the gambler's ruin probability); while the probability for a critical branching process to survive at least k generations is $2/(\text{Var}[X]k)$ [13, 14]. This result, together with the matching of the first two cumulants of the total population of a subcritical branching process[15] with those of the corresponding Brownian excursion, leads to the identifications:

$$v = \frac{E[X] - 1}{\text{Var}[X]}, \quad D = \frac{1}{\text{Var}[X]}, \quad x_0 = \frac{2}{\text{Var}[X]}, \quad (2)$$

where $E[X] = (z-1)p$ and $\text{Var}[X] = (z-1)p(1-p)$. This is in agreement with the known rescaling of the so-called contour process [16]. In addition, the system is conditioned to percolate, so that the associated walk must reach height $h(j) = L$ for some $0 < j < 2S$ before returning to the origin for the first time at $i = 2S$. In summary, the large L limit of the walk $(i, h(i))$ corresponds to a Brownian excursion of maximum height L , drift v and diffusion constant D . The (random) duration T of this Brownian excursion equals $2S$ by construction. Proper convergence of a positive walk to a Brownian ex-

curtion is formally only reached after rescaling lengths appropriately, but we will only take this step at the very end of our calculations, once the scaling of S with L , as a function of p , has been calculated. Formal proofs of convergence in similar setups can be found in the related literature on *continuum random trees*. For instance, Aldous[12] shows that if certain families of trees are constrained to have a fixed large number of nodes, then its associated Harris walks converge to the standard Brownian excursion of length 1; while Le Gall[17] also considers constraints related to the maximum height. In both cases it is shown that, under certain mild conditions, all families of trees (i.e., all lattices) lead to the same universal results.

We have thus set up a mapping between percolation in a finite Bethe lattice and a Brownian excursion with a reflecting boundary, and explicitly related their parameters, Eq. (2). In what follows, we will compute the distribution of T in the Brownian excursion setting, and show that the results agree, as expected, with simulations of percolation in a Bethe lattice.

BROWNIAN EXCURSION

The Brownian excursion of fixed height L corresponding to our problem can be decomposed into two paths (see Fig. 2): (i) a first path from the origin $x(0) = 0$ at time $t = 0$ to the boundary $x(T_1) = L$, taking time T_1 , and (ii) a return path from the reflecting boundary $x(T_1) = L$ to the origin $x(T) = 0$, taking time $T_2 = T - T_1$. Note that the first path involves a conditional exit at $x = L$ such that that boundary is effectively absorbing for $t < T_1$. Readers experienced with Brownian processes will recognize T_1 as the hitting time to reach L of a 3d Bessel process[5].

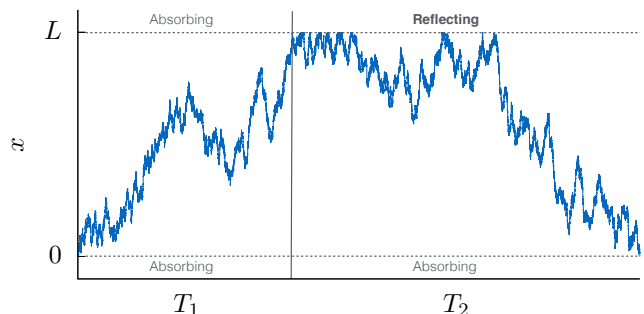


FIG. 2. A Brownian excursion, conditioned to reach $x(t) = L$ for some $0 < t < T = T_1 + T_2$, decomposed into (i) the path from $x = 0$ that reaches $x = L$ for the first time, (ii) the return path from $x = L$ back to $x = 0$.

Diffusion equation with drift

The distributions of T_1 and T_2 are most conveniently calculated by working with the Laplace transform of the diffusion equation with drift:

$$[-D\partial_{xx} + v\partial_x + \lambda]\hat{\phi}(x, \lambda) = \delta(x - x_0) \quad (3)$$

where $\phi(x, \lambda)$ is the Laplace transform with respect to time of the density function with initial condition $x(0) = x_0$. According to the usual recipe[18], this equation is solved with combinations of exponentials respecting the boundary conditions, and fluxes are computed at the relevant boundaries. These calculations furnish the Laplace transformed first passage densities $j_1(\lambda)$ and $j_2(\lambda)$ associated with T_1 and T_2 .

First path: A + A

To calculate T_1 for the first path of the decomposition, absorbing boundaries are required at both at $x = 0$ and $x = L$,

$$\hat{\phi}_{AA}(0, \lambda) = 0 \quad (4)$$

$$\hat{\phi}_{AA}(L, \lambda) = 0, \quad (5)$$

together with a conditional exit $x = L$. $\hat{\phi}_{AA}$ satisfying the absorbing boundaries takes the form

$$\hat{\phi}_{AA}(x, \lambda) = \frac{2e^{v(x-x_0)/2D}}{\rho \sinh(\rho L/2D)} \sinh(\rho x_0/2D) \sinh(\rho(L-x)/2D) \quad (6)$$

where $\rho := \sqrt{v^2 + 4\lambda D}$. The conditional exit probability at $x = L$ is[18]

$$\text{Prob}(\text{exit at } x = L) = \frac{1 - e^{-vx_0/D}}{1 - e^{-Lv/D}} \quad (7)$$

so that the flux at the boundary $x = L$ is

$$j_1(\lambda) = \lim_{x_0 \rightarrow 0} \frac{-D\hat{\phi}'_{AA}(L, \lambda)}{\text{Prob}(\text{exit at } x = L)} \quad (8)$$

$$= \frac{\rho \sinh(Lv/2D)}{v \sinh(\rho L/2D)} \quad (9)$$

Return path: A + R

To calculate T_2 for the second path of the decomposition, an absorbing boundary is required at $x = 0$

$$\hat{\phi}_{AR}(0, \lambda) = 0 \quad (10)$$

and a Dirac δ -pulse is injected at $x = L$ as an initial condition, which in Laplace space is equivalent to the boundary condition

$$-v\hat{\phi}_{AR}(L, \lambda) + D\hat{\phi}'_{AR}(L, \lambda) = 1 \quad (11)$$

$\hat{\phi}_{AR}$ satisfying these boundary conditions is

$$\hat{\phi}_{AR}(x, \lambda) = \frac{2e^{v(x-L)/2D} \sinh(\rho x/2D)}{\rho \cosh(\rho L/2D) - v \sinh(\rho L/2D)}, \quad (12)$$

so that the flux at the boundary $x = 0$ is

$$j_2(\lambda) = D\hat{\phi}'_{AR}(0, \lambda) \quad (13)$$

$$= \frac{\rho e^{-Lv/2D}}{\rho \cosh(\rho L/2D) - v \sinh(\rho L/2D)}. \quad (14)$$

Entire excursion

The total time to go from the origin to the boundary at $x = L$, and then from the boundary back to the origin is $T = T_1 + T_2$. Since T_1 and T_2 are independent, the Laplace transform of the density of the total time T is the product

$$\begin{aligned} j(\lambda) &= j_1(\lambda) \cdot j_2(\lambda) \\ &= \frac{\rho^2 \sinh(Lv/2D)}{v \sinh(\rho L/2D)} \times \\ &\quad \frac{e^{-Lv/2D}}{\rho \cosh(\rho L/2D) - v \sinh(\rho L/2D)} \end{aligned} \quad (15)$$

Critical case

Critical percolation with $p = p_c = 1/(z-1)$ maps to Brownian motion with zero drift. Taking the limit $v \rightarrow 0$ in Eq.(15),

$$j_1^{(c)}(\lambda) = \frac{L\sqrt{\lambda/D}}{\sinh(L\sqrt{\lambda/D})} \quad (16)$$

$$j_2^{(c)}(\lambda) = \frac{1}{\cosh(L\sqrt{\lambda/D})} \quad (17)$$

$$j^{(c)}(\lambda) = \frac{2L\sqrt{\lambda/D}}{\sinh(2L\sqrt{\lambda/D})}. \quad (18)$$

Interestingly, $j_1^{(c)}$ and $j^{(c)}$ are Laplace transforms of the same KS distribution on different scales. $j_2^{(c)}$ is also the Laplace transform of many observables of Brownian motions, see[5] for a review.

General case

For general choices of v (or, equivalently, p), the full solution in Eq. (15) cannot be inverted explicitly, and one must resort to numerical inversion. Fig 3 shows a family of percolating cluster size distributions, all rescaled by their means. The distributions are independent of coordination number z as expected, but clearly depend

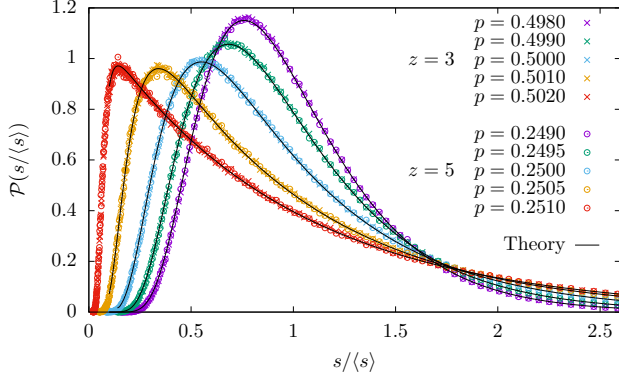


FIG. 3. The rescaled distribution of the percolating cluster size, for different values of z and p and fixed system size $L = 10^3$. Simulations (colored symbols) agree with theory (black line), which is obtained via numerical inversion of Eq.(15).

on p . For $p \neq p_c$ (and fixed L), the distributions can be thought of as flowing away from the non-trivial fixed point, represented by the KS distribution (middle, blue curve).

To better understand these flows towards trivial fixed points (in the language of renormalization-group (RG)) for $p \rightarrow 0$ and $p \rightarrow 1$, it is convenient to standardize the distributions to zero mean and unit standard deviation. Since Eq. (15) is the moment generating function for T , these cumulants can be constructed as usual by differentiating once (first moment) or twice (second moment) with respect to λ , and taking $\lambda \rightarrow 0$. For the mean $\mu := \langle T \rangle$ we find

$$\mu \frac{v^2}{D} = \mathcal{F}(Lv/D), \quad (19)$$

where the scaling function

$$\mathcal{F}(x) = e^x - 3 + \frac{2x}{e^x - 1} \quad (20)$$

behaves as e^x for $x \gg 1$, x^2 for $x \rightarrow 0$, and x for $x \ll -1$. For fixed v and D , the scaling behavior of μ is therefore

$$\mu = \begin{cases} \frac{D}{v^2} e^{Lv/D}, & L \gg 1, v > 0 \\ \frac{2}{3D} L^2, & L \gg 1, v = 0 \\ \frac{2}{|v|} L, & L \gg 1, v < 0. \end{cases} \quad (21)$$

Expressions can likewise be obtained for the standard deviation $\sigma := \sqrt{\langle T^2 \rangle - \langle T \rangle^2}$ (not shown).

Given μ and σ , we arrive at a standardized characteristic function $\mathcal{G}(k; Lv/D)$ from Eq. (15) by (i) replacing $\lambda \rightarrow -ik$, to change from Laplace to Fourier space, (ii) multiplying by $\exp(-ik\mu)$ to translate the distribution to zero mean, and (iii) replacing $k \rightarrow k/\sigma$, to scale the distribution to unit standard deviation. Note that the ratio μ/σ is a function of Lv/D alone. Thus, $\mathcal{G}(k; Lv/D)$

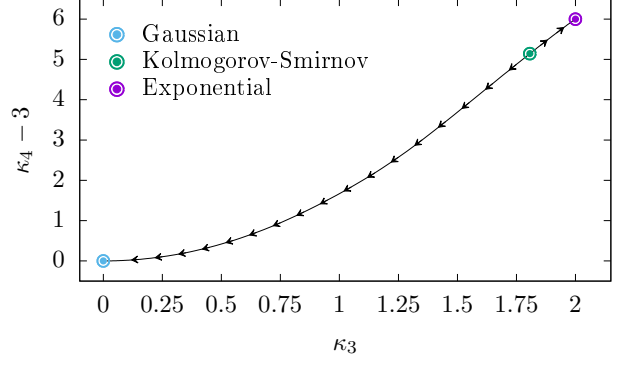


FIG. 4. Centered skewness κ_3 and kurtosis $\kappa_4 - 3$ for the standardized solution described by $\mathcal{G}(k; Lv)$. Solid circles from left to right correspond to the fixed points: Gaussian (subcritical), KS (critical), Exponential (supercritical).

describes a family of probability distributions parameterized by Lv/D .

As a result of this standardization, all distributions $\mathcal{G}(k; Lv/D)$ have zero mean and unit standard deviation. However, higher cumulants κ_n will depend on Lv/D . These differences can be visualized by plotting the skewness κ_3 versus the kurtosis $\kappa_4 - 3$, as depicted in Fig. 4. Three regimes are particularly noteworthy: (i) $Lv/D \rightarrow -\infty$ (subcritical fixed point), (ii) $Lv/D \rightarrow 0$ (critical fixed point), (iii) $Lv/D \rightarrow \infty$ (supercritical fixed point). With the help of Mathematica, we can obtain full expressions for $(\kappa_3, \kappa_4 - 3)$ as a function of Lv/D . In the three above regimes these reduce to (i) $(2, 6)$, (ii) $(4\sqrt{10}/7, 36/7)$, (iii) $(0, 0)$, corresponding to exponential, KS and Gaussian distributions, respectively. These three fixed points are marked in Fig. 4 as solid circles. A heuristic explanation for the sub and supercritical limits is as follows: the overall behavior of T is dominated by T_2 (T_1 is insensitive to the sign of v , which can be seen from the fact that (6) is an even function of v). When $v < 0$, T_2 is approximately the ballistic travel time from $x = L$ to $x = 0$, with a Gaussian correction coming from diffusion. When $v > 0$, T_2 is dominated by the rare event that the motion escapes the drift that keeps returning it to the reflecting boundary at $x = L$. This rare event process is consistent with exponential tails.

We briefly comment on how our method differs from that of Botet and Płoszajczak[8], the full details of which are presented in[9]. In their method, a recursion is set up between successive generations of the Bethe lattice, encoding the statistical weights of configurations conditioned to percolate. The recursion relation is then Laplace transformed and rescaled by mean percolating cluster size, to yield expressions which are then analyzed asymptotically in the limit of large system size for the critical case $p = p_c$ [19]. In contrast, our method makes

use of a passage from continuum trees to Brownian excursions at the starting point of the calculation, such that the parameters of the Bethe lattice are reincorporated into the diffusion constant and drift of the resulting motion. This makes the subsequent analysis arguably more intuitive, and Eq.(15) gives the entire cluster size distribution for all regimes from the solution of a diffusion problem with drift. Our approach should be universal across different types of trees, in the sense that the fixed point distributions (suitably rescaled) do not depend on underlying microscopic details.

CONCLUSION

We have calculated the distribution of the size of the percolating cluster on a tree by interpreting the problem as a type of Brownian excursion. In this way, we give an intuitive explanation for the coincidence in distribution (first noted in[8]) with other observables associated with Brownian motions[5]. The analysis can be extended off criticality by adding a drift term to the associated diffusion equation. The resulting flows in the space of distributions can be captured by tracking the skewness and kurtosis. Our exact calculation makes an investigation of the various regimes possible in full detail. We expect that such mappings can be used to investigate further properties of branching processes.

The authors would like to thank Álvaro Corral, Gregory Schehr and Zoe Budrikis for useful discussions.

-
- [1] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, UK, 1996).
 - [2] F. C. Auluck and D. S. Kothari, Proc. Cambridge Philos. Soc. **42**, 272 (1946).
 - [3] M. L. Mehta, *Random Matrices* (Academic Press, 2004).
 - [4] J. L. Doob, Ann. Math. Statist. **20**, 393 (1946).
 - [5] P. Biane, J. Pitman, and M. Yor, Bull. Amer. Math. Soc. (N.S.) **38**, 435 (2001).
 - [6] G. S. Watson, Biometrika **48** (1961).
 - [7] G. Foltin, K. Oerding, Z. Rácz, R. L. Workman, and R. K. P. Zia, Phys. Rev. E **50**, R639 (1994).
 - [8] R. Botet and M. Płoszajczak, Phys. Rev. Lett. **95**, 185702 (2005).
 - [9] R. Botet, J. Phys.: Conf. Ser. **297**, 012005 (2011).
 - [10] S. N. Majumdar, Curr. Sci. **89**, 2076 (2005).
 - [11] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1994).
 - [12] D. Aldous, Ann. Probab. **21**, 248 (1993).
 - [13] T. E. Harris, *The Theory of Branching Processes (Grundlehren der mathematischen Wissenschaften)* (Springer, 1963).
 - [14] K. B. Athreya and P. E. Ney, *Branching Processes (Dover Books on Mathematics)* (Dover Publications, 2004).
 - [15] Y. Suhov and M. Kelbert, *Probability and Statistics by Example: Volume 1, Basic Probability and Statistics*, 2nd ed. (Cambridge University Press, 2014).
 - [16] J. Bennes and G. Kersting, J. Theoret. Probab. **13**, 777 (2000).
 - [17] J. F. L. Gall, Probab. Surveys **2**, 245 (2005).
 - [18] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
 - [19] Off criticality, Botet reports (but does not show) Gaussian statistics, which we find only in the subcritical case.